



-71-

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Schlessinger, Joseph
Sap, Jan M.
- (ii) TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
PHOSPHATASE-ALPHA
- (iii) NUMBER OF SEQUENCES: 14
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: PENNIE & EDMONDS
 - (B) STREET: 1155 AVENUE OF THE AMERICAS
 - (C) CITY: NEW YORK
 - (D) STATE: NEW YORK
 - (E) COUNTRY: U.S.A.
 - (F) ZIP: 10036
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/015,985
 - (B) FILING DATE: 10-FEB-1993
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Coruzzi, Laura A.
 - (B) REGISTRATION NUMBER: 30,742
 - (C) REFERENCE/DOCKET NUMBER: 7683-020
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (212) 790-9090
 - (B) TELEFAX: (212) 869-9741/8864
 - (C) TELEX: 66141 PENNIE

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 802 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met Asp Ser Trp Phe Ile Leu Val Leu Leu Gly Ser Gly Leu Ile Cys
1 5 10 15
Val Ser Ala Asn Asn Ala Thr Thr Val Ala Pro Ser Val Gly Ile Thr
20 25 30
Arg Leu Ile Asn Ser Ser Thr Ala Glu Pro Val Lys Glu Glu Ala Lys
35 40 45
Thr Ser Asn Pro Thr Ser Ser Leu Thr Ser Leu Ser Val Ala Pro Thr
50 55 60

Phe Ser Pro Asn Ile Thr Leu Gly Pro Thr Tyr Leu Thr Thr Val Asn
 65 70 75 80
 Ser Ser Asp Ser Asp Asn Gly Thr Thr Arg Thr Ala Ser Thr Asn Ser
 85 90 95
 Ile Gly Ile Thr Ile Ser Pro Asn Gly Thr Trp Leu Pro Asp Asn Gln
 100 105 110
 Phe Thr Asp Ala Arg Thr Glu Pro Trp Glu Gly Asn Ser Ser Thr Ala
 115 120 125
 Ala Thr Thr Pro Glu Thr Phe Pro Pro Ser Gly Asn Ser Asp Ser Lys
 130 135 140
 Asp Arg Arg Asp Glu Thr Pro Ile Ile Ala Val Met Val Ala Leu Ser
 145 150 155 160
 Ser Leu Leu Val Ile Val Phe Ile Ile Ile Val Leu Tyr Met Leu Arg
 165 170 175
 Phe Lys Lys Tyr Lys Gln Ala Gly Ser His Ser Asn Ser Phe Arg Leu
 180 185 190
 Ser Asn Gly Arg Thr Glu Asp Val Glu Pro Gln Ser Val Pro Leu Leu
 195 200 205
 Ala Arg Ser Pro Ser Thr Asn Arg Lys Tyr Pro Pro Leu Pro Val Asp
 210 215 220
 Lys Leu Glu Glu Glu Ile Asn Arg Arg Met Ala Asp Asp Asn Lys Leu
 225 230 235 240
 Phe Arg Glu Glu Phe Asn Ala Leu Pro Ala Cys Pro Ile Gln Ala Thr
 245 250 255
 Cys Glu Ala Ala Ser Lys Glu Glu Asn Lys Glu Lys Asn Arg Tyr Val
 260 265 270
 Asn Ile Leu Pro Tyr Asp His Ser Arg Val His Leu Thr Pro Val Glu
 275 280 285
 Gly Val Pro Asp Ser Asp Tyr Ile Asn Ala Ser Phe Ile Asn Gly Tyr
 290 295 300
 Gln Glu Lys Asn Lys Phe Ile Ala Ala Gln Gly Pro Lys Glu Glu Thr
 305 310 315 320
 Val Asn Asp Phe Trp Arg Met Ile Trp Glu Gln Asn Thr Ala Thr Ile
 325 330 335
 Val Met Val Thr Asn Leu Lys Glu Arg Lys Glu Cys Lys Cys Ala Gln
 340 345 350
 Tyr Trp Pro Asp Gln Gly Cys Trp Thr Tyr Gly Asn Ile Arg Val Ser
 355 360 365
 Val Glu Asp Val Thr Val Leu Val Asp Tyr Thr Val Arg Lys Phe Cys
 370 375 380
 Ile Gln Gln Val Gly Asp Met Thr Asn Arg Lys Pro Gln Arg Leu Ile
 385 390 395 400
 Thr Gln Phe His Phe Thr Ser Trp Pro Asp Phe Gly Val Pro Phe Thr
 405 410 415
 Pro Ile Gly Met Leu Lys Phe Leu Lys Lys Val Lys Ala Cys Asn Pro

Tyr Lys Val Val Gln Glu Tyr Ile Asp Ala Phe Ser Asp Tyr Ala Asn
785 790 795 800

Phe Lys

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2409 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ATGGATTCCT GGTTCATTCT TGTTCGCTC GGCAGTGGTC TGATATGTGT CAGTGCCAAC	60
AATGCTACCA CAGTTGCACC TTCTGTAGGA ATTACAAGAT TAATTAAGTC ATCAACGGCA	120
GAACCAAGTA AAGAAGAGGC CAAAAGTTCA AATCCAAGTT CTTCACTAAC TTCTCTTTCT	180
GTGGCACCAG CATTGAGCCC AAATATAACT CTGGGACCCA CCTATTTAAC CACTGTCAAT	240
TCTTCAGACT CTGACAATGG GACCACAAGA ACAGCAAGCA CCAATTCTAT AGGCATTACA	300
ATTTACACCA ATGGAACGTG GCTTCCAGAT AACCAGTTCA CGGATGCCAG AACAGAACCC	360
TGGGAGGGGA ATTCCAGCAC CGCAGCAACC ACTCCAGAAA CTTTCCCTCC TTCAGGTAAT	420
TCTGACTCGA AGGACAGAAG AGATGAGACA CCAATTATTG CGGTGATGGT GGCCCTGTCC	480
TCTCTGCTAG TGATCGTGTT TATTATCATA GTTTTGTACA TGTTAAGGTT TAAGAAATAC	540
AAGCAAGCTG GGAGCCATTC CAATTCTTTC CGCTTATCCA ACGGCCGCAC TGAGGATGTG	600
GAGCCCCAGA GTGTGCCACT TCTGGCCAGA TCCCCAAGCA CCAACAGGAA ATACCCACCC	660
CTGCCCCGTG ACAAGCTGGA AGAGGAAATT AACCAGGAGAA TGGCAGACGA CAATAAGCTC	720
TTCAAGGAGG AATTCAACGC TCTCCCTGCA TGTCTATCC AGGCCACCTG TGAGGCTGCT	780
TCCAAGGAGG AAAACAAGGA AAAAATCGA TATGTAAACA TCTTGCCTTA TGACCACTCT	840
AGAGTCCACC TGACACCGGT TGAAGGGGTT CCAGATTCTG ATTACATCAA TGCTTCATTC	900
ATCAACGGTT ACCAAGAAAA GAACAAATTC ATTGCTGCAC AAGGACCAA AGAAGAAACG	960
GTGAATGATT TCTGGCGGAT GATCTGGGAA CAAAACACAG CCACCATCGT CATGGTTACC	1020
AACCTGAAGG AGAGAAAGGA GTGCAAGTGC GCCCAGTACT GGCCAGACCA AGGCTGCTGG	1080
ACCTATGGGA ATATTGGGGT GTCTGTAGAG GATGTGACTG TCCTGGTGGA CTACACAGTA	1140
CGGAAGTTCT GCATCCAGCA GGTGGGCGAC ATGACCAACA GAAAGCCACA GCGCCTCATC	1200
ACTCAGTTCC ACTTTACCAG CTGGCCAGAC TTTGGGGTGC CTTTACCCC GATCGGCATG	1260
CTCAAGTTCC TCAAGAAGGT GAAGGCCTGT AACCCTCAGT ATGCAGGGGC CATCGTGGTC	1320
CACTGCAGTG CAGGTGTAGG GCGTACAGGT ACCTTTGTCT TCATTGATGC CATGCTGGAC	1380
ATGATGCATA CAGAACGGAA GGTGGACGTG TATGGCTTTG TGAGCCCGAT CCGGGCACAG	1440
CGCTGCCAGA TGGTGCAAAC CGATATGCAG TATGTCTTCA TATACCAAGC CTTTCTGGAG	1500

CATTATCTCT ATGGAGATAC AGAACTGGAA GTGACCTCTC TAGAAACCCA CCTGCAGAAA 1560
 ATTTACAACA AAATCCCAGG GACCAGCAAC AATGGATTAG AGGAGGAGTT TAAGAAGTTA 1620
 ACATCAATCA AAATCCAGAA TGACAAGATG CGGACTGGAA ACCTTCCAGC CAACATGAAG 1680
 AAGAACCGTG TTTTACAGAT CATTCCATAT GAATTCAACA GAGTGATCAT TCCAGTTAAG 1740
 CGGGGCGAAG AGAATACAGA CTATGTGAAC GCATCCTTTA TTGATGGCTA CCGGCAGAAG 1800
 GACTCCTATA TCGCCAGOCA GGGCCCTCTT CTCCACACAA TTGAGGACTT CTGGCGAATG 1860
 ATCTGGGAGT GGAAATCCTG CTCTATCGTG ATGCTAACAG AACTGGAGGA GAGAGGCCAG 1920
 GAGAAGTGTG CCCAGTACTG GCCATCTGAT GGACTGGTGT CCTATGGAGA TATTACAGTG 1980
 GAACTGAAGA AGGAGGAGGA ATGTGAGAGC TACACCGTCC GAGACCTCCT GGTCACCAAC 2040
 ACCAGGGAGA ATAAGAGCCG GCAGATCCGG CAGTTCCACT TCCATGGCTG GCCTGAAGTG 2100
 GGCATCCCCA GTGACGGAAA GGGCATGATC AGCATCATCG CCGCCGTGCA GAAGCAGCAG 2160
 CAGCAGTCAG GGAACCACCC CATCACCGTG CACTGCAGCG CCGGGGCAGG AAGGACGGGG 2220
 ACCTTCTGTG CCCTGAGCAC CGTCCTGGAG CGTGTGAAAG CAGAGGGGAT TTTGGATGTC 2280
 TTCCAGACTG TCAAGAGCCT GCGGCTACAG AGGCCACACA TGGTCCAGAC ACTGGAACAG 2340
 TATGAGTTCT GCTACAAGGT GGTGCAGGAG TATATTGATG CATTCTCAGA TTATGCCAAC 2400
 TTCAAGTAA 2409

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 793 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Asp Ser Trp Phe Ile Leu Val Leu Phe Gly Ser Gly Leu Ile His
 1 5 10 15
 Val Ser Ala Asn Asn Ala Thr Thr Val Ser Pro Ser Leu Gly Thr Thr
 20 25 30
 Arg Leu Ile Lys Thr Ser Thr Thr Glu Leu Ala Lys Glu Glu Asn Lys
 35 40 45
 Thr Ser Asn Ser Thr Ser Ser Val Ile Ser Leu Ser Val Ala Pro Thr
 50 55 60
 Phe Ser Pro Asn Leu Thr Leu Glu Pro Thr Tyr Val Thr Thr Val Asn
 65 70 75 80
 Ser Ser His Ser Asp Asn Gly Thr Arg Arg Ala Ala Ser Thr Glu Ser
 85 90 95
 Gly Gly Thr Thr Ile Ser Pro Asn Gly Ser Trp Leu Ile Glu Asn Gln
 100 105 110
 Phe Thr Asp Ala Ile Thr Glu Pro Trp Glu Gly Asn Ser Ser Thr Ala
 115 120 125

Ala Thr Thr Pro Glu Thr Phe Pro Pro Ala Asp Glu Thr Pro Ile Ile
130 135 140

Ala Val Met Val Ala Leu Ser Ser Leu Leu Val Ile Val Phe Ile Ile
145 150 155 160

Ile Val Leu Tyr Met Leu Arg Phe Lys Lys Tyr Lys Gln Ala Gly Ser
165 170 175

His Ser Asn Ser Phe Arg Leu Ser Asn Gly Arg Thr Glu Asp Val Glu
180 185 190

Pro Gln Ser Val Pro Leu Leu Ala Arg Ser Pro Ser Thr Asn Arg Lys
195 200 205

Tyr Pro Pro Leu Pro Val Asp Lys Leu Glu Glu Glu Ile Asn Arg Arg
210 215 220

Met Ala Asp Asp Asn Lys Leu Phe Arg Glu Glu Phe Asn Ala Leu Pro
225 230 235 240

Ala Cys Pro Ile Gln Ala Thr Cys Glu Ala Ala Ser Lys Glu Glu Asn
245 250 255

Lys Glu Lys Asn Arg Tyr Val Asn Ile Leu Pro Tyr Asp His Ser Arg
260 265 270

Val His Leu Thr Pro Val Glu Gly Val Pro Asp Ser Asp Tyr Ile Asn
275 280 285

Ala Ser Phe Ile Asn Gly Tyr Gln Glu Lys Asn Lys Phe Ile Ala Ala
290 295 300

Gln Gly Pro Lys Glu Glu Thr Val Asn Asp Phe Trp Arg Met Ile Trp
305 310 315 320

Glu Gln Asn Thr Ala Thr Ile Val Met Val Thr Asn Leu Lys Glu Arg
325 330 335

Lys Glu Cys Lys Cys Ala Gln Tyr Trp Pro Asp Gln Gly Cys Trp Thr
340 345 350

Tyr Gly Asn Val Arg Val Ser Val Glu Asp Val Thr Val Leu Val Asp
355 360 365

Tyr Thr Val Arg Lys Phe Ser Ile Gln Gln Val Gly Asp Val Thr Asn
370 375 380

Arg Lys Pro Gln Arg Leu Ile Thr Gln Phe His Phe Thr Ser Trp Pro
385 390 395 400

Asp Phe Gly Val Pro Phe Thr Pro Ile Gly Met Leu Lys Phe Leu Lys
405 410 415

Lys Val Lys Ala Cys Asn Pro Gln Tyr Ala Gly Ala Ile Val Val His
420 425 430

Cys Ser Ala Gly Val Gly Arg Thr Gly Thr Phe Val Val Ile Asp Ala
435 440 445

Met Leu Asp Met Met His Ser Glu Arg Lys Val Asp Val Tyr Gly Phe
450 455 460

Val Ser Arg Ile Arg Ala Gln Arg Cys Gln Met Val Gln Thr Asp Met
465 470 475 480

Gln Tyr Val Phe Ile Tyr Gln Ala Leu Leu Glu His Tyr Leu Tyr Gly

485										490										495										
Asp	Thr	Glu	Leu	Glu	Val	Thr	Ser	Leu	Glu	Thr	His	Leu	Gln	Lys	Ile															
			500					505					510																	
Tyr	Asn	Lys	Ile	Pro	Gly	Thr	Ser	Asn	Asn	Gly	Leu	Glu	Glu	Glu	Phe															
		515					520					525																		
Lys	Lys	Leu	Thr	Ser	Ile	Lys	Ile	Gln	Asn	Asp	Lys	Met	Arg	Thr	Gly															
	530					535					540																			
Asn	Leu	Pro	Ala	Asn	Met	Lys	Lys	Asn	Arg	Val	Leu	Gln	Ile	Ile	Pro															
545					550				555						560															
Tyr	Glu	Phe	Asn	Arg	Val	Ile	Ile	Pro	Val	Lys	Arg	Gly	Glu	Glu	Asn															
				565				570						575																
Thr	Asp	Tyr	Val	Asn	Ala	Ser	Phe	Ile	Asp	Gly	Tyr	Arg	Gln	Lys	Asp															
			580					585					590																	
Ser	Tyr	Ile	Ala	Ser	Gln	Gly	Pro	Leu	Leu	His	Thr	Ile	Glu	Asp	Phe															
		595				600						605																		
Trp	Arg	Met	Ile	Trp	Glu	Trp	Lys	Ser	Cys	Ser	Ile	Val	Met	Leu	Thr															
	610					615					620																			
Glu	Leu	Glu	Glu	Arg	Gly	Gln	Glu	Lys	Cys	Ala	Gln	Tyr	Trp	Pro	Ser															
625					630					635					640															
Asp	Gly	Leu	Val	Ser	Tyr	Gly	Asp	Ile	Thr	Val	Glu	Leu	Lys	Lys	Glu															
				645				650						655																
Glu	Glu	Cys	Glu	Ser	Tyr	Thr	Val	Arg	Asp	Leu	Leu	Val	Thr	Asn	Thr															
			660					665					670																	
Arg	Glu	Asn	Lys	Ser	Arg	Gln	Ile	Arg	Gln	Phe	His	Phe	His	Gly	Trp															
		675				680						685																		
Pro	Glu	Val	Gly	Ile	Pro	Ser	Asp	Gly	Lys	Gly	Met	Ile	Asn	Ile	Ile															
	690				695					700																				
Ala	Ala	Val	Gln	Lys	Gln	Gln	Gln	Gln	Ser	Gly	Asn	His	Pro	Ile	Thr															
705					710					715					720															
Val	His	Cys	Ser	Ala	Gly	Ala	Gly	Arg	Thr	Gly	Thr	Phe	Cys	Ala	Leu															
				725				730						735																
Ser	Thr	Val	Leu	Glu	Arg	Val	Lys	Ala	Glu	Gly	Ile	Leu	Asp	Val	Phe															
			740				745						750																	
Gln	Thr	Val	Lys	Ser	Leu	Arg	Leu	Gln	Arg	Pro	His	Met	Val	Gln	Thr															
		755					760					765																		
Leu	Glu	Gln	Tyr	Glu	Phe	Cys	Tyr	Lys	Val	Val	Gln	Glu	Tyr	Ile	Asp															
	770					775					780																			
Ala	Phe	Ser	Asp	Tyr	Ala	Asn	Phe	Lys																						
785					790																									

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2872 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GAATTCCGGC GAGTGAGGCG CTGACAGGGA CTCGCGGGGG CATCTTGCAC AGACCCCTGG	60
ACCACGCCGC CATCGCAGCC TCCAGCCCAG TCCTCTCTCT GCCGCTTCTC CTCGCCATGG	120
AGGCCGCCGA CCGCCGTCCG CGGGCTTCGA GCAGCGGACC GGGCCGGGCT GACCCCATGT	180
GGGCCGAGAG CCCGGTCCTG AGGCGGAGCT GCCGTGCGCG TCCCCGCGG TCCCGCCCCA	240
GCGCCGGGCT CCGTCAGCAT GGATTCTTGG TTCATTCTTG TCCTGTTTGG CAGTGGTCTA	300
ATACATGTTA GTGCCAACAA TGCTACTACA GTTTCACCTT CTTTAGGAAC GACAAGATTA	360
ATTAAACAT CAACAACAGA ATTGGCTAAG GAAGAGAATA AAACCTCAAA TTCAACCTCT	420
TCAGTAATTT CTCTTTCTGT GGCACCAACA TTCAGCCCAA ACCTGACTCT GGAGCCCACC	480
TATGTGACTA CTGTTAATTC TTCACACTCT GACAATGGGA CCAGGAGGGC AGCCAGCAGC	540
GAATCTGGAG GCACTACCAT TTCCCGAAC GGAAGCTGGC TTATTGAGAA CCAGTTCACG	600
GATGCCATAA CAGAACCCTG GGAGGGGAAC TCCAGCACTG CAGCAACCAC TCCAGAAACC	660
TTCCCCCGG CAGATGAGAC ACCAATTATT GCGGTGATGG TGGCCCTGTC CTCTCTGCTA	720
GTAATCGTGT TTATTATCAT AGTTCGTGAC ATGTTAAGGT TTAAGAAATA CAAGCAAGCT	780
GGGAGTCATT CCAACTCTTT CCGCCTGTCA AATGGCCGCA CGGAGGATGT GGAGCCCCAA	840
AGTGTACCAC TTCTGGCCAG GTCCCCGAGC ACCAACAGGA AGTACCCACC ACTGCCTGTG	900
GACAAGCTGG AAGAGGAGAT TAACCGGAGA ATGGCTGATG ACAATAAGCT CTTCAGAGAA	960
GAATTC AACG CTCTCCCTGC TTGTCCTATC CAGGCCACCT GTGAGGCTGC CTCCAAGGAA	1020
GAAACAAGG AAAAAAACCG CTATGTAAAC ATCCTGCCCT ATGACCACTC TAGAGTGCAC	1080
CTGACACCTG TTGAAGGGGT CCCAGATTCT GATTACATCA ACGCTTCATT CATTAAATGGC	1140
TACCAGGAAA AGAACAAATT CATCGCTGCA CAAGGACCAA AAGAAGAAAC AGTGAATGAC	1200
TTCTGGAGAA TGATATGGGA AAAAAACACA GCTACTATTG TCATGGTGAC CAACCTGAAG	1260
GAGAGAAAGG AGTGTAATG TGCCCAATAC TGGCCAGACC AAGGCTGCTG GACCTATGGG	1320
AATGTCCGTG TGTCTGTGCA GGATGTGACT GTTCTGGTGG ACTACACAGT ACGGAAATTC	1380
TCGATCCAGC AGGTGGGCGA CGTGACCAAC AGGAAACCAC AGCGCCTCAT CACTCAGTTC	1440
CACTTCACCA GCTGGCCAGA CTTTGGGGTG CCTTTCACCC CAATTGGCAT GCTCAAGTTC	1500
CTCAAGAAGG TGAAGGCCTG TAACCCTCAG TACGCAGGGG CTATCGTGGT CCACTGCAGT	1560
GCAGGTGTAG GCGGCACTGG CACCTTTGTT GTCATCGATG CCATGCTGGA CATGATGCAT	1620
TCGGAGCGCA AAGTGGATGT ATATGGGTTT GTGAGCCGGA TCCGGGCCCCA GCGCTGCCAG	1680
ATGGTACAGA CAGACATGCA GTACGTCTTC ATATACCAGG CCCTTCTGGA GCATTATCTG	1740
TATGGGGACA CAGAACTGGA AGTGACTTCT CTAGAAACCC ACCTACAAA AATTATAAC	1800
AAGATCCCAG GGAAGTACAA CAACGGGTTA GAGGAGGAGT TTAAGAAATT AACTTCAATC	1860
AAAATCCAGA ATGACAAGAT GCGCACGGGA AACCTTCCAG CCAACATGAA GAAGAACCGG	1920

GTTTTACAGA TCATTCCATA TGAATTTAAC AGAGTGATCA TTCCAGTCAA ACGAGGCGAA 1980
 GAGAACACAG ACTATGTGAA CGCATCCTTC ATTGATGGAT ACCGGCAGAA AGACTCCTAC 2040
 ATTGCCAGCC AGGGCCCTCT TCTCCACACG ATTGAGGACT TCTGGCGAAT GATCTGGGAG 2100
 TGGAAGTCCT GTTCTATCGT AATGCTGACA GAACTGGAAG AGAGAGGCCA GGAGAAGTGT 2160
 GCCCAGTACT GGCCATCTGA TGGCCTGGTG TCCTACGGAG ACATCACAGT TGAGCTGAAG 2220
 AAGGAGGAGG AATGTGAAAG CTACACTGTC CGAGACCTCC TGGTCACCAA CACCAGGGAG 2280
 AACAAAGAGTC GGCAAATCCG GCAGTTCCAC TTCCACGGCT GGCCTGAGGT GGGCATCCCC 2340
 AGCGACGGCA AGGGCATGAT CAACATCATT GCAGCAGTGC AGAAGCAGCA GCAGCAGTGC 2400
 GGGAAACCATC CCATCACTGT GCACTGCAGT GCCGGGGCAG GACGGACAGG AACCTTCTGT 2460
 GCCTTGAGCA CAGTCCTGGA ACGTGTGAAA GCAGAAGGAA TTTTAGATGT CTTCCAAACT 2520
 GTCAAGAGCC TGC GGCTGCA GAGGCCACAC ATGGTCCAGA CACTGGAACA GTATGAATTC 2580
 TGCTACAAGG TGGTACAGGA ATACATTGAC GCCTTTTTCAG ATTATGCCAA CTTCAAGTGA 2640
 CAGGTGACAA GGCCACAGA CAGGAGAATT GCCTTTAATA TTTTGTAATA TTCTGTTTTT 2700
 GTTAATATAC CAAAATTGT ATATATCTTA TAACTGTTTT AGAAATGGCA CATAGGCTTC 2760
 TATTACCTGT TAGATGGAGA TTTTGTATGT AAATGTGTTA GCACTGATAG TCCTTTTCCA 2820
 GTGTTTTATT GGGAAATTAA TAGTGTGATA TTTGGGTGA TATAATGAAT TC 2872

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 235 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Asn Gln Asn Lys Asn Arg Tyr Val Asp Ile Leu Pro Tyr Asp Tyr Asn
 1 5 10 15
 Arg Val Glu Leu Ser Glu Ile Asn Gly Asp Ala Gly Ser Asn Tyr Ile
 20 25 30
 Asn Ala Ser Tyr Ile Asp Gly Phe Lys Glu Pro Arg Lys Tyr Ile Ala
 35 40 45
 Ala Gln Gly Pro Arg Asp Glu Thr Val Asp Asp Phe Trp Arg Met Ile
 50 55 60
 Trp Glu Gln Lys Ala Thr Val Ile Val Met Val Thr Arg Cys Glu Glu
 65 70 75 80
 Gly Asn Arg Asn Lys Cys Ala Glu Tyr Trp Pro Ser Met Glu Glu Gly
 85 90 95
 Thr Arg Ala Phe Gly Asp Val Val Val Lys Ile Asn Gln His Lys Arg
 100 105 110
 Cys Pro Asp Tyr Ile Ile Gln Lys Leu Asn Ile Val Asn Lys Lys Glu

115	120	125
Lys Ala Thr Gly Arg Glu Val Thr His Ile Gln Phe Thr Ser Trp Pro		
130	135	140
Asp His Gly Val Pro Glu Asp Pro His Leu Leu Leu Lys Leu Arg Arg		
145	150	155
Arg Val Asn Ala Phe Ser Asn Phe Phe Ser Gly Pro Ile Val Val His		
165	170	175
Cys Ser Ala Gly Val Gly Arg Thr Gly Thr Tyr Ile Gly Ile Asp Ala		
180	185	190
Met Leu Glu Gly Leu Glu Ala Glu Asn Lys Val Asp Val Tyr Gly Tyr		
195	200	205
Val Val Lys Leu Arg Arg Gln Arg Cys Leu Met Val Gln Val Glu Ala		
210	215	220
Gln Tyr Ile Leu Ile His Gln Ala Leu Val Glu		
225	230	235

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 236 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Asn Lys Glu Lys Asn Arg Tyr Val Asn Ile Leu Pro Tyr Asp His Ser		
1	5	10
Arg Val His Leu Thr Pro Val Glu Gly Val Pro Asp Ser Asp Tyr Ile		
20	25	30
Asn Ala Ser Phe Ile Asn Gly Tyr Gln Glu Lys Asn Lys Phe Ile Ala		
35	40	45
Ala Gln Gly Pro Lys Glu Glu Thr Val Asn Asp Phe Trp Arg Met Ile		
50	55	60
Trp Glu Gln Asn Thr Ala Thr Ile Val Met Val Thr Asn Leu Lys Glu		
65	70	75
Arg Lys Glu Cys Lys Cys Ala Gln Tyr Trp Pro Asp Gln Gly Glu Trp		
85	90	95
Thr Tyr Gly Asn Ile Arg Val Ser Val Glu Asp Val Thr Val Leu Val		
100	105	110
Asp Tyr Thr Val Arg Lys Phe Cys Ile Gln Gln Val Gly Asp Met Thr		
115	120	125
Asn Arg Lys Pro Gln Arg Leu Ile Thr Gln Phe His Phe Thr Ser Trp		
130	135	140
Pro Asp Phe Gly Val Pro Phe Thr Pro Ile Gly Met Leu Lys Phe Leu		
145	150	155
Lys Lys Val Lys Ala Cys Asn Pro Gln Tyr Ala Gly Ala Ile Val Val		

	165		170		175
His Cys Ser Ala Gly Val Gly Arg Thr Gly Thr Phe Val Val Ile Asp					
	180		185		190
Ala Met Leu Asp Met Met His Thr Glu Arg Lys Val Asp Val Tyr Gly					
	195		200		205
Phe Val Ser Arg Ile Arg Ala Gln Arg Cys Gln Met Val Gln Thr Asp					
	210		215		220
Met Gln Tyr Val Phe Ile Tyr Gln Ala Leu Leu Glu					
	225		230		235

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 242 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Asn Lys His Lys Asn Arg Tyr Ile Asn Ile Val Ala Tyr Asp His Ser					
1	5		10		15
Arg Val Lys Leu Ala Gln Leu Ala Glu Lys Asp Gly Lys Leu Thr Asp					
	20		25		30
Tyr Ile Asn Ala Asn Tyr Val Asp Gly Tyr Asn Arg Pro Lys Ala Tyr					
	35		40		45
Ile Ala Ala Gln Gly Pro Leu Lys Ser Thr Ala Glu Asp Phe Trp Arg					
	50		55		60
Met Ile Trp Glu His Asn Val Glu Val Ile Val Met Ile Thr Asn Leu					
	65		70		75
Val Glu Lys Gly Arg Arg Lys Cys Asp Gln Tyr Trp Pro Ala Asp Gly					
	85		90		95
Ser Glu Glu Tyr Gly Asn Phe Leu Val Thr Gln Lys Ser Val Gln Val					
	100		105		110
Leu Ala Tyr Tyr Thr Val Arg Asn Phe Thr Leu Arg Asn Thr Lys Ile					
	115		120		125
Lys Lys Gly Ser Gln Lys Gly Arg Pro Ser Gly Arg Val Val Thr Gln					
	130		135		140
Tyr His Tyr Thr Gln Trp Pro Asp Met Gly Val Pro Glu Tyr Ser Leu					
	145		150		155
Pro Val Leu Thr Phe Val Arg Lys Ala Ala Tyr Ala Lys Arg His Ala					
	165		170		175
Val Gly Pro Val Val Val His Cys Ser Ala Gly Val Gly Arg Thr Gly					
	180		185		190
Thr Tyr Ile Val Leu Asp Ser Met Leu Gln Gln Ile Gln His Glu Gly					
	195		200		205
Thr Val Asn Ile Phe Gly Phe Leu Lys His Ile Arg Ser Gln Arg Asn					



(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 245 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

[illegible]

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 248 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(ix) FEATURE:

- (A) NAME/KEY: Modified-sites
- (B) LOCATION: 1..248
- (D) OTHER INFORMATION: /label= Xaa
/note= "For the Consensus Sequence, Xaa = Lack of Consensus"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```

Asn Lys His Lys Asn Arg Tyr Xaa Asn Ile Leu Xaa Tyr Asp His Ser
1      5      10
Arg Val Lys Leu Xaa Xaa Leu Xaa Xaa Lys Xaa Xaa Lys Xaa Ser Asp
20     25     30
Tyr Ile Asn Ala Xaa Tyr Xaa Asp Gly Tyr Asn Glu Pro Lys Xaa Tyr
35     40     45
Ile Ala Ala Gln Gly Pro Leu Lys Xaa Thr Val Glu Asp Phe Trp Arg
50     55     60
Met Ile Trp Glu Gln Asn Thr Xaa Val Ile Val Met Xaa Thr Asn Leu
65     70     75     80
Val Glu Lys Gly Arg Arg Lys Cys Xaa Gln Tyr Trp Pro Xaa Xaa Gly
85     90     95
Ser Glu Xaa Tyr Gly Asn Ile Xaa Val Thr Val Lys Xaa Val Xaa Val
100    105    110
Leu Ala Xaa Xaa Asp Tyr Thr Val Arg Lys Phe Xaa Xaa Arg Asn Thr
115    120    125
Lys Ile Xaa Lys Xaa Gly Xaa Lys Xaa Xaa Xaa Lys Gly Arg Xaa Xaa
130    135    140
Gly Arg Val Val Thr Gln Tyr His Xaa Thr Xaa Trp Pro Asp Met Gly
145    150    155    160
Val Pro Glu Tyr Pro Leu Pro Val Leu Xaa Phe Val Arg Xaa Val Xaa
165    170    175
Ala Ala Xaa Xaa Xaa Xaa Xaa Gly Pro Xaa Val Val His Cys Ser Ala
180    185    190
Gly Val Gly Arg Thr Gly Thr Tyr Ile Val Ile Asp Xaa Met Leu Gln
195    200    205
Gln Ile Xaa Xaa Glu Xaa Xaa Val Xaa Val Tyr Gly Phe Xaa Lys His
210    215    220
Ile Arg Xaa Gln Arg Xaa Tyr Xaa Val Gln Thr Glu Glu Gln Tyr Xaa
225    230    235    240
Phe Ile His Xaa Ala Leu Xaa Glu
245

```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 260 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```

Asn Lys Ser Lys Asn Arg Asn Ser Asn Val Ile Pro Tyr Asp Tyr Asn
1      5      10      15
Arg Val Pro Leu Lys His Glu Leu Glu Met Ser Lys Glu Ser Glu His
      20      25      30
Asp Ser Asp Glu Ser Ser Asp Asp Asp Ser Asp Ser Glu Glu Pro Ser
      35      40      45
Lys Tyr Ile Asn Ala Ser Phe Ile Met Ser Tyr Trp Lys Pro Glu Val
      50      55      60
Met Ile Ala Ala Gln Gly Pro Leu Lys Glu Thr Ile Gly Asp Phe Trp
      65      70      75      80
Gln Met Ile Phe Gln Arg Lys Val Lys Val Ile Val Met Leu Thr Glu
      85      90      95
Leu Lys His Gly Asp Gln Glu Ile Cys Ala Gln Tyr Trp Gly Glu Gly
      100      105      110
Lys Gln Thr Tyr Gly Asp Ile Glu Val Asp Leu Lys Asp Thr Asp Lys
      115      120      125
Ser Ser Thr Tyr Thr Leu Arg Val Phe Glu Leu Arg His Ser Lys Arg
      130      135      140
Lys Asp Ser Arg Thr Val Tyr Gln Tyr Gln Tyr Thr Asn Trp Ser Val
      145      150      155      160
Glu Gln Leu Pro Ala Glu Pro Lys Glu Leu Ile Ser Met Ile Gln Val
      165      170      175
Val Lys Gln Lys Leu Pro Gln Lys Asn Ser Ser Glu Gly Asn Lys His
      180      185      190
His Lys Ser Thr Pro Leu Leu Ile His Cys Arg Asp Gly Ser Gln Gln
      195      200      205
Thr Gly Ile Phe Cys Ala Leu Leu Asn Leu Leu Glu Ser Ala Glu Thr
      210      215      220
Glu Glu Val Val Asp Ile Phe Gln Val Val Lys Ala Leu Arg Lys Ala
      225      230      235      240
Arg Pro Gly Met Val Ser Thr Phe Glu Gln Tyr Gln Phe Leu Tyr Asp
      245      250      255
Val Ile Ala Ser
      260

```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 233 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```

Asn Met Lys Lys Asn Arg Val Leu Gln Ile Ile Pro Tyr Glu Phe Asn
1           5           10           15
Arg Val Ile Ile Pro Val Lys Arg Gly Glu Glu Asn Thr Asp Tyr Val
20           25           30
Asn Ala Ser Phe Ile Asp Gly Tyr Arg Gln Lys Asp Ser Tyr Ile Ala
35           40           45
Ser Gln Gly Pro Leu Leu His Thr Ile Glu Asp Phe Trp Arg Met Ile
50           55           60
Trp Glu Trp Lys Ser Cys Ser Ile Val Met Leu Thr Glu Leu Glu Glu
65           70           75           80
Arg Gly Gln Glu Lys Cys Ala Gln Tyr Trp Pro Ser Asp Gly Leu Val
85           90           95
Ser Tyr Gly Asp Ile Thr Val Glu Leu Lys Lys Glu Glu Glu Cys Glu
100          105          110
Ser Tyr Thr Val Arg Asp Leu Leu Val Thr Asn Thr Arg Glu Asn Lys
115          120          125
Ser Arg Gln Ile Arg Gln Phe His Phe His Gly Trp Pro Glu Val Gly
130          135          140
Ile Pro Ser Asp Gly Lys Gly Met Ile Ser Ile Ile Ala Ala Val Gln
145          150          155          160
Lys Gln Gln Gln Gln Ser Gly Asn His Pro Ile Thr Val His Cys Ser
165          170          175
Ala Gly Ala Gly Arg Thr Gly Thr Phe Cys Ala Leu Ser Thr Val Leu
180          185          190
Glu Arg Val Lys Ala Glu Gly Ile Leu Asp Val Phe Gln Thr Val Lys
195          200          205
Ser Leu Ala Leu Gln Arg Pro His Met Val Gln Thr Leu Glu Gln Tyr
210          215          220
Glu Phe Cys Tyr Lys Val Val Gln Glu
225          230

```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 234 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```

Asn Arg Glu Lys Asn Arg Thr Ser Ser Ile Ile Pro Val Glu Arg Ser
1           5           10           15
Arg Val Gly Ile Ser Ser Leu Ser Gly Glu Gly Thr Asp Tyr Ile Asn
20           25           30
Ala Ser Tyr Ile Met Gly Tyr Tyr Gln Ser Asn Glu Phe Ile Ile Thr
35           40           45
Gln His Pro Leu Leu His Thr Ile Lys Asp Phe Trp Arg Met Ile Trp
50           55           60
Asp His Asn Ala Gln Leu Val Val Met Ile Pro Asp Gly Gln Asn Met
65           70           75           80
Ala Glu Asp Glu Phe Val Tyr Trp Pro Asn Lys Asp Glu Pro Ile Asn
85           90           95
Cys Glu Ser Phe Lys Val Thr Leu Met Ala Glu Glu His Lys Cys Leu
100          105          110
Ser Asn Glu Glu Lys Leu Ile Ile Gln Asp Phe Ile Leu Glu Ala Thr
115          120          125
Gln Asp Asp Tyr Val Leu Glu Val Arg His Phe Gln Cys Pro Lys Trp
130          135          140
Pro Asn Pro Asp Ser Pro Ile Ser Lys Thr Phe Glu Leu Ile Ser Val
145          150          155          160
Ile Lys Glu Glu Ala Ala Asn Arg Asp Gly Pro Met Ile Val His Asp
165          170          175
Glu His Gly Gly Val Thr Ala Gly Thr Phe Cys Ala Leu Thr Thr Leu
180          185          190
Met His Gln Leu Glu Lys Glu Asn Ser Val Asp Val Tyr Gln Val Ala
195          200          205
Lys Met Ile Asn Leu Met Arg Pro Gly Val Phe Ala Asp Ile Glu Gln
210          215          220
Tyr Gln Phe Leu Tyr Lys Val Ile Leu Ser
225          230

```

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 235 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```

Asn Lys Glu Lys Asn Arg Asn Ser Ser Val Val Pro Ser Glu Arg Ala
1           5           10           15
Arg Val Gly Leu Ala Pro Leu Pro Gly Met Lys Gly Thr Asp Tyr Ile
20           25           30
Asn Ala Ser Tyr Ile Met Gly Tyr Tyr Arg Ser Asn Glu Phe Ile Ile
35           40           45

```



```

Thr Gln His Pro Leu Pro His Thr Thr Lys Asp Phe Trp Arg Met Ile
 50                      55-          60

Trp Asp His Asn Ala Gln Ile Ile Val Met Leu Pro Asp Asn Gln Ser
65                      70          75          80

Leu Ala Glu Asp Glu Phe Val Tyr Trp Pro Ser Arg Glu Glu Ser Met
                      85          90          95

Asn Cys Glu Ala Phe Thr Val Thr Leu Ile Ser Lys Asp Arg Leu Cys
          100                      105          110

Leu Ser Asn Glu Glu Gln Ile Ile Ile His Asp Phe Ile Leu Glu Ala
          115                      120          125

Thr Gln Asp Asp Tyr Val Leu Glu Val Arg His Phe Gln Cys Pro Lys
130                      135          140

Trp Pro Asn Pro Asp Ala Pro Ile Ser Ser Thr Phe Glu Leu Ile Asn
145                      150          155          160

Val Ile Lys Glu Glu Ala Leu Thr Arg Asp Gly Pro Thr Ile Val His
          165                      170          175

Asp Glu Tyr Gly Ala Val Ser Ala Gly Met Leu Cys Ala Leu Thr Thr
          180          185          190

Leu Ser Gln Gln Leu Glu Asn Glu Asn Ala Val Asp Val Phe Gln Val
          195          200          205

Ala Lys Met Ile Asn Leu Met Arg Pro Gly Val Phe Thr Asp Ile Glu
210                      215          220

Gln Tyr Gln Phe Ile Tyr Lys Ala Arg Leu Ser
225                      230          235

```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 280 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(ix) FEATURE:

- (A) NAME/KEY: Modified-sites
- (B) LOCATION: 1..280
- (D) OTHER INFORMATION: /label= Xaa
/note= "For the Consensus Sequence, Xaa = Lack of Consensus"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```

Asn Lys Glu Lys Asn Arg Asn Ser Ser Xaa Ile Pro Tyr Glu Arg Asn
 1                      5          10          15

Arg Val Gly Xaa Xaa Xaa Leu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
20                      25          30

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Gly Glu Glu Gly Thr
35                      40          45

Asp Tyr Ile Asn Ala Ser Xaa Ile Met Gly Tyr Tyr Gln Ser Asn Glu
50                      55          60

```

Phe Ile Xaa Thr Gln Xaa Pro Leu Leu His Thr Ile Lys Asp Phe Trp
 65 70 75 80
 Arg Met Ile Trp Asp His Xaa Asn Ala Gln Ile Val Met Leu Xaa Xaa
 85 90 95
 Xaa Gln Xaa Xaa Ala Glu Xaa Glu Xaa Gln Tyr Trp Pro Ser Xaa
 100 105 110
 Gly Xaa Xaa Xaa Tyr Gly Asp Xaa Xaa Val Xaa Leu Lys Xaa Xaa Xaa
 115 120 125
 Asn Cys Glu Ser Xaa Thr Val Thr Xaa Xaa Xaa Glu Xaa Arg Xaa Cys
 130 135 140
 Leu Ser Asn Glu Xaa Arg Xaa Ile Ile Gln Asp Phe Ile Leu Glu Ala
 145 150 155 160
 Thr Gln Asp Asp Tyr Val Leu Glu Val Arg His Phe Gln Cys Pro Lys
 165 170 175
 Trp Pro Asn Pro Asp Xaa Pro Ile Ser Xaa Thr Xaa Glu Leu Ile Ser
 180 185 190
 Val Ile Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Gln Lys Xaa Glu Glu Ala
 195 200 205
 Xaa Asn Arg Xaa Xaa Xaa Asp Gly Pro Xaa Ile Val His Xaa Glu Xaa
 210 215 220
 Gly Ala Val Xaa Xaa Gly Thr Phe Cys Ala Leu Thr Thr Leu Leu Glu
 225 230 235 240
 Gln Leu Glu Xaa Glu Asn Xaa Val Asp Val Phe Gln Val Xaa Lys Met
 245 250 255
 Xaa Asn Leu Met Arg Pro Gly Xaa Xaa Xaa Xaa Ile Glu Gln Tyr Gln
 260 265 270
 Phe Leu Tyr Lys Val Ile Leu Ser
 275 280